Running Head: LEGISLATING TECHNOLOGY

MetaTech Consulting, Inc.

Whitepaper

Legislation of a Technology

Jim Thomas

April 2, 2004

Abstract

It is the purpose of this paper to provide a treatment of the laws and policies surrounding *Total Information Awareness* – a United States Department of Defense program concerned with *data mining*. A high-level description of the program with a more detailed discussion of its technical approach begins the paper. This is followed by a presentation of the arguments levied by the opponents of the program. The laws most germane to the program are addressed prior to a summary of the program's current state. A statement of concern that continued development of a single technology has been terminated as a result of concerns for personal privacy concludes the paper.

Table of Contents

Abstract 2
Table of Contents 3
Legislation of a Technology 4
Program Description 4
Technical Approach 5
  Technologies 6
  Strategies 7
Arguments 9
  Technology 9
  Privacy 10
Related Laws and Policies 12
Public Law 108-7 13
State of the Program 15
Conclusions 15
References 16
Figure 1 18
Figure 2 19
Table 1 20

Legislation of a Technology

It is the purpose of this paper to provide a treatment of the laws and policies surrounding *Total Information Awareness* – a United States Department of Defense program concerned with *data mining*. A high-level description of the program with a more detailed discussion of its technical approach begins the paper. Having established a sufficient context for further discussions, the paper continues with a presentation of the arguments levied by the opponents of the program. These include doubts that the proposed technology – data mining – is a viable solution and concern that the program will result in atrocious invasions of personal privacy. Public Law 108-7, the piece of legislation that limited the funding for and deployment of the TIA program technologies, is reviewed to provide an understanding of the concluding remarks. A statement of concern that continued development of a single technology has been terminated as a result of concerns for personal privacy concludes the paper.

Program Description

The Terrorism Information Awareness (TIA) program, conceived as the Total Information Awareness program, was established under Information Awareness Office (IAO) – headed by retired Vice Admiral John Poindexter – within the Defense Advanced Research Projects Agency (DARPA) of the Department of Defense (DoD) in 2002. DARPA, as characterized by Poindexter (2003), has conducted high-risk, high-payoff for the DoD since its inception in 1958. The IAO was created in January of 2002 to focus research on the challenges of counterterrorism – the need for such an office was made evident by the attacks on United States soil by terrorists on September 11, 2001 (*Report to Congress*, 2003).

Mack (2002), in documenting the systems description, stated the objective of the system was to provide the United States Department of Defense with a system of end-to-end capabilities for analysts and decision-makers. The four key elements characterizing the program were a) employment of an innovative architecture based on the Open Grid Service Architecture (OGSA), b) use of a rapid turn-around experimentation processes focusing on real data and addressing real operational issues, c) using a technology refresh process that enables the introduction of the latest technologies in real operational settings, and d) with an infrastructure development and tool hardening process to facilitate the transition of the best of breed tools to uses in the operational environments.

Poindexter amplified the objectives of the program and made clear the system would be able to detect, classify, identify, and track terrorists so that decision-makers can understand and interrupt the terrorist plans (Poindexter, 2002). He went on to describe TIA as the activity that would integrate the functionality provided by several other programs rather than a program to develop or implement the functionality anew. As illustrated by Belasco (2003), congress authorized funding for TIA as an integration activity separate from the technology programs linked to TIA. This substantiates the assertion presented by Poindexter. In response to concern from zealous advocates for civil liberties, the moniker of the program was changed in 2003 through its stated objectives remained unchanged.

Technical Approach

It is necessary to posses at least a fundamental understanding of the technologies related to TIA in order to appreciate the resistance and concerns of its opponents. The following few paragraphs are intended to provide a context for the remaining sections of

the paper by providing a concise description of the technologies involved and the strategy TIA intends to deploy them.

Technologies

Stevens (2003) enumerates the TIA core technology types as a) data search and pattern recognition, b) machine translation of languages, and c) advanced collaborative and decision support. These areas will be addressed in turn presently.

Data search and pattern recognition. This set of technologies is generally referred to as *data mining*. This term has justifiable meaning to the generalist and specialist alike. To the generalist, the term refers to any number of techniques by which he is able to examine data – generally located in some form of database(s) – to gain awareness of hidden facts, obscure details, or possibly previously unknown relationships existing in varied data sets. To some, a series of searches on the internet satisfies their definition of the term. To the specialist, data mining connotes a far more discrete set of activities including sophisticated algorithms executed against prepared data sets existing in specialized data structures. The end objective of generalist and specialist are quite similar though the rigor employed differs dramatically. Seifert (2003) succinctly characterized this set of technologies in saying that it "involves the use of data analysis tools to discover preciously unknown, valid patterns and relationships in large data sets." According to Belasco (2003), funding of TIA efforts in data mining was $29.2 million in fiscal year (FY) 2001, $38.2 million in FY2002, and $53.0 million in FY2003. This is the single most contentious TIA-related technology. Arguments both for and against the use of this technology by the government will be presented within the appropriate section of this paper as appropriate.

Machine translation of languages. The logical grouping of technologies is intended to make evident to English-speaking users of the system information that is persisted in foreign languages and also for converting speak to text (Poindexter, 2002). These initiatives had a relatively constant level of approximately $36 million annually between FY2001 and FY2003 (Belasco, 2003).

Advanced collaboration and decision support. Stevens (2003) characterized technologies of this type as those supporting war-gaming simulations, collaborative reasoning processes. The intended purpose of these technologies, in the TIA context, is to enable the high-level decision makers to anticipate, recognize, preempt, and react to terrorism in both operational and training conditions. Belasco (2003) reported the funding of TIA efforts of this type was $14.4 million in fiscal year (FY) 2001, $23.5 million in FY2002, and $39.5 million in FY2003.

Strategies

The challenges confronting the TIA program are not founded solely in technology; there are legal issues at play as well. The following discussion presents two of the key strategies the program has selected together with the associated challenge.

Architecture. The TIA functional processing flow (Figure 1) offered by Mack (2002) emphasizes the dependence on storing and sharing information to the viability of the architecture. Though the depiction suggests a central repository is intended, such a conclusion is likely inaccurate. A single view of a system, particularly such an abstract model, cannot convey all system details. Stanley (2003) provides a different depiction of TIA (Figure 2) – a chart revealing the data sources and data flow of the system. This view suggests that the system will utilize many different physical databases, some

purpose built for the program and others leveraged from the source system of the data of interest. The number, location, and ownership of the database(s) are less points of technical significance than of legal significance. Further discussion of this issue will be deferred until the appropriate section below.

Data mining. The tools and algorithms of data mining have continued to mature within the scientific, technical, and business communities over the past two decades. One familiar with the technology may argue that it is far too mature to be of interest to DARPA. It is true that the concepts, principles, and mechanics are sufficiently mature that in-and-of themselves, they are of little interest to DARPA. Why it is appropriate for DARPA to concern itself with data mining is the unproven, innovative approach the technology is being applied. Typical data mining strategies require an engineering activity to focus on limited data sets that are meticulously prepared and stored in specially designed constructs known as multidimensional databases. Frequently, the data to be mined is stored in a relational database such as that in an OnLine Transaction Processing (OLTP) system common to many businesses. To be effectively mined, a more meaningful construct commonly based on facts and dimensions of the data will be built. The data would then be transformed or otherwise manipulated as it is moved from the source system to the mining environment. In the new, optimized structure, the mining algorithms can effectively churn through vast quantities of data to discover trends and relationships that would otherwise be invisible.

The mining strategy envisioned for TIA deviates from the aforementioned techniques in several interesting ways. The number of entities (e.g. facts and dimensions) considered relevant in most mining activities is quite small – the fewer entities, the more

probable the algorithms will be successful in discovering a pattern. The TIA approach will strive to mine across very large numbers of entities (Figure 2). This is possible because TIA will first construct hypotheses about relationships – known as templates, models, scenarios, or patterns - and look for evidence of those relationships in the data. (Tether, 2003).

Arguments

The challenges levied by TIA opponents are founded on two broad arguments: a) the underlying technologies are insufficient to constitute a viable solution, and b) the concept of operations for the system represents a fundamental infringement to personal privacy rights. The following paragraphs provide discussion of these two broad arguments.

Technology

There are a number of technological challenges that must be overcome for TIA to provide an effective tool in the war against terrorism. If the technologies involved were sufficiently mature and if the implementation strategies were proven, there would not be a need for DARPA to undertake this development. The technology challenges are real and they are significant. This is precisely why further development and experimentation should be undertaken. The most significant technological obstacles, together with how they impact the concept of operations, are discussed presently.

Accuracy of source data. TIA is to function by exploiting data persisted in many disparate repositories. While some of the data is to be in ingested data into purpose-built data stores, much of the data will simply be accessed from its source systems. This point of distinction is significant in that ingesting data provides an opportunity to cleanse the

content to improve its overall quality - duplicate data can be resolved, conflicting data can be reconciled, inaccurate data can be corrected. Data cleansing is a non-trivial process that is imprecise and cannot assure perfect results. Data mining requires highly accurate content to produce quality results. The confidence one may attribute to conclusions formed through data mining is proportional to the quality of the data that was mined.

False positives. Data mining activities performed on commercial data is often undertaken with well understood and proven interrelationships between the various data. While much of the TIA data is the same that may be considered in other commercial mining activities, the interrelationships of interest are not so well understood. In addition to simple purchase transactions, travel details, etc., TIA will strive to discover the relationships between those well understood data types as the behavior of terrorist. Through codifying the behavior of known terrorist, and by speculating about their likely behavior under given circumstances, TIA hopes to predict terrorist activity and to assist the authorities in intervening early in the planning phase. Poorly understood behavior of the terrorists, poorly understood relationships between the data sets of interest, and poor quality of the source data will compound the probability of misinterpreting behavior and actions. This will result in missing indicators of actual terrorist activity and in classifying benign behavior of citizens as terrorist-related. While either result is detrimental, it is the later that most offends opponents of TIA.

Privacy

The most substantial challenges to TIA are based on the personal privacy rights of United States citizens. Our transaction economy generates an abundance of detailed data

on everything from what movies we rent to our eating habits to our individual recreational activities to the medication prescribed by our physicians. These data exist in myriad databases owned, maintained, and used by countless merchants, government agencies, care givers, and service providers. We generally trust that the information maintained about us is needed by the data-owner to provide us products or services for which we transact business with them.

The government has long been subject to constraints limited what information they can maintain on individuals. Such protection was not extended to citizens in the private sector. During the last decade, we found that our personal information that had been collected by one service provider was being given or sold to other service providers. Businesses were formed for the sole purpose of trading in personal information (e.g. demographic data, spending data, and information in the public record). Within the past few years, legislation such as the E-Government Act of 2002 has been enacted that requires those entities maintaining information about us to make explicit how they intend to use the information and if they intend to make it available to make it available to others (Dempsey, 2003). The following paragraphs treat the prominent personal privacy issues germane to TIA.

Types of data. The TIA concept of operations identifies the following three broad classes of data supporting the antiterrorism campaign: a) intelligence data, b) authentication biometric data, and c) transactional data. Within the transactional class are the most controversial data sources – financial, educational, travel, medical, housing, and communications. As the government is generally prevented from maintaining these

forms of data itself, it must be obtained from the private sector. This is possible either by first obtaining permission (e.g. subpoena) or by simply purchasing access to the data.

Improper use of data. Opponents of TIA have voiced concern in how the vast quantities of personal data might be misused by those who are entrusted with access to it. It is not difficult to imagine how one with access to such a comprehensive resource, one who may be suspicious of his or her significant other, might use the information at hand to satisfy compelling curiosity. Likewise, one may give in to the temptation of exploiting vulnerabilities of an acquaintance via blackmail. Real and imagined examples of both of these forms of impropriety have been dramatized by the media since the earliest days of information management systems. The new concern is that so much information would be vulnerable to exploitation. This elevates another closely related concern: such a tremendous repository of information would be an ideal candidate for hacking and other forms of intrusion attacks. If one were to gain access to the TIA, the potential for damage to (e.g. personal character, financial, etc.) would be limitless. Security solutions are too immature to offset the risk of bringing such much personal information within a single system (B. Simmons & E. Spafford, personal communication, January 23, 2003).

Related Laws and Policies

In the report to congress on the privacy issues germane to TIA, Stevens (2003) characterizes the comprehensiveness of federal statues pertaining to the privacy of information as follows:

> Federal laws protect government, credit, communications, education, bank, cable, video, motor vehicle, health, telecommunications, children's, and financial information. These laws generally carve out exceptions for

> the disclosure of personally identifiable information to law enforcement officials, and authorize access to personal information through use of search warrants, subpoenas, and court orders. (p. 5)

The report to congress also makes note of the fact that there are no blanket prohibitions on access to publicly available information by the federal government. This point gives deference to the notion of *access*. Unless explicitly addressed by targeted legislation, it is understood that the federal government may legally access any publicly available information. Table 1 contains a taxonomy of laws outlined in the report. Those interested in a thorough treatment of each of these laws are encouraged to read the report in total.

Public Law 108-7

Since the announcement of the TIA program, and the ensuing controversy, lawmakers have been compelled to intervene in its execution through restrictive legislation. On 3 February, 2003 a Joint Resolution from the House of Representatives, making further continuing appropriations for fiscal year 2003, was signed into law as Public Law 108-7 (H. Res. 108-7, 2003). Section 111(a) of that Act explicitly limited the use of funds for the program then named Total Information Awareness beyond 90 days of its enactment. The legislation provided two exceptions allowing for continued spending on the program. The principle exception, detailed in Section 111(b) of the Act, was a written report jointly by the Secretary of Defense, the Attorney General, and the Director of Central Intelligence submitted to Congress.

The report was required to contain (a) a detailed explanation of the actual and intended use of funds under the program, (b) the schedule for proposed research and

development, and (c) target dates for the deployment of each project and activity of the program. The report was also required to address the viability and efficacy of program in "providing practically valuable predictive assessments of the plans, intentions, or capabilities of terrorists or terrorist groups" (H. Res. 108-7, 2003, p. 1119). Additionally, the Act required the joint authors to address in the report an assessment of the likely impact a program such as TIA would have on privacy and civil liberties as well and to provide a listing of the laws and regulations governing the information collected by the TIA together with a description of any changes that would be need for the program to function as intended.

Section 111(c) of the Act made explicit the limitations on the deployment of the TIA. Under the Act, the no portion of the TIA is allowed to be deployed without the specific authorization with appropriate appropriations by Congress. A key exception to the stated limitation allowed for deployment of the system in support of "lawful military operations conducted outside of the United States …. [or] lawful foreign intelligence activities conducted wholly overseas, or wholly against non-United States persons" (H. Res. 108-7, 2003, p. 1121).

On May 20, 2003 the report authored on behalf of the Secretary of Defense, the Attorney General, and the Director of Central Intelligence titled *Report to Congress regarding the Terrorism Information Awareness Program* was submitted to Congress. This 26 page report – reflecting the newly adopted name of the program - fully satisfied its requirements. It attempted to convey the expectation that TIA would simply develop the capability to mine data for a given purpose and that the deployment strategy of the system would be left to the lawmakers as they deemed necessary and appropriate.

State of the Program

As reported in the Washington Post (Hudson, 2003), the final defense spending bill for 2004 closed down the DARPA Information Awareness Office and terminates data mining efforts within the TIA program. DARPA will be allowed to continue research that does not involve data mining and the National Foreign Intelligence Program will not be restricted from using those technology tools developed under TIA for foreign intelligence purposes overseas.

Conclusions

The Information Assurance Office and the Total Information Awareness program were created in reaction to the terrorist events of September 11, 2001. The program was to bring together under a single concept of operations, and through an integration activity, several disparate research and development activities already underway at the DARPA. The intent of the program was to detect and prevent future terrorist activities by examining data from numerous sources to ascertain if they matched hypothesized terrorist planning patterns. A groundswell of privacy rights activism, fronted by the American Civil Liberties Union and the Electronic Privacy Information Center, effectively influenced congressional leaders and brought the program to a hasty end.

The author finds it disturbing not that a promising tool was removed from our nation's arsenal for the war on terrorism, but that a single technology could be so vilified. The arguments against TIA manifestly targeted data mining as a technology rather than focusing on the sensitivity of the data sources identified for exploitation. The legislation enacted prevents further development of a technology that, if used judiciously, is arguably our single best hope in preventing future strikes against our homeland.

References

Belasco, A. (2003). Total information awareness programs: Funding, composition, and oversight issues. *Congressional Research Service Report RL31786.* The Library of Congress. March 21, 2003.

Defense Advanced Research Projects Agency, Information Awareness Office (2003). *Report to Congress regarding the Terrorist Information Program: In response to Consolidated Appropriations Resolution, 2003, Pub. L. No. 108-7, Division M, Section 111(b).* 20 May, 2003.

H. Res. 108-7, 108th Cong., Cong. Rec. 1118 (2003) (enacted).

Hudson, A. (2003). Congress kills data-mining computer program. *The Washington Times.* September 26, 2003. Retrieved March 21, 2004 from http://www.washingtimes.com/functions/print.php?StoryID=20030925-115156-7992r.

Mack, G. (2002). *Total Information Awareness Program (TIA) system description document, v 1.1.* DARPA Information Awareness Office. July 19, 2002.

Poindexter. J. M. (2002). *Overview of the Information Awareness Office.* DARPATech Conference, Anaheim, CA, August 2, 2002. Retrieved March 28, 2004 from http://www.fas.org/irp/agency/dod/poindexter.html.

Poindexter, J. M. (2003). Finding the face of terror in data. *The New York Times.* September 10, 2003. Retrieved March 27, 2004 from http://www.nytimes.com/2003/10/opinion/1OPINION.html.

Seifert, J. W. (2003). Data mining: An overview. *Congressional Research Service Report RL31798.* The Library of Congress. March 21, 2003.

Stanley, J. (2003). *Total information compliance: The TIA's burden under the Wyden amendment.* American Civil Liberties Union Technology and Liberty Program. May 19, 2003.

Stevens, G. M. (2003). Privacy: Total information awareness programs and related information access, collection, and protection laws. *Congressional Research Service Report RL31730.* The Library of Congress. March 21, 2003.

Tether, T. (2003). Statement submitted to the United Stated House of Representatives Subcommittee on Technology, Information Policy, Intergovernmental Relations and the Subcommittee on Governmental Reform. Retrieved April 6, 2004 from http://frwebgate.access.gpo.gov/cgi-bin/useftp.cgi?IPaddress=162.140.64.21&filename=90399.pdf&directory=/diskb/wais/data/108_house_hearings

The Defense of Privacy Act and privacy in the hands of the government: *Hearing before the House of Representatives Committee on the Judiciary Subcommittee on Commercial and Administrative Law and Subcommittee on the Constitution,* 108th Cong.(2003). Retrieved March 27, 2004 from http://www.cdt.org/testimony/030722dempsey.shtml.

Figure 1

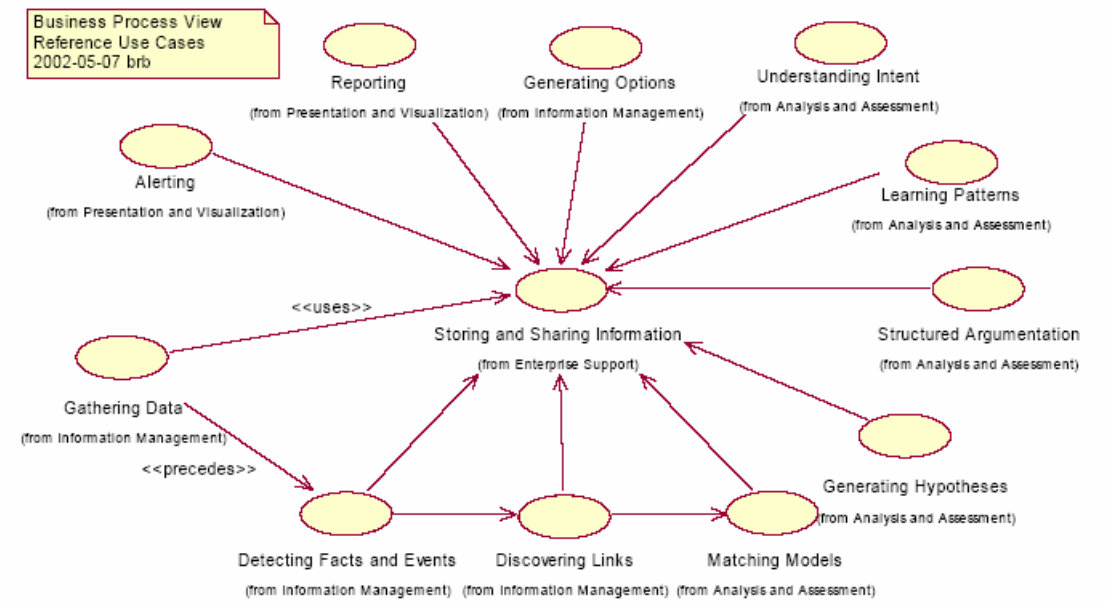

Figure 1. TIA Functional Processing Flow (Mack, 2002, p. 16)

Figure 2

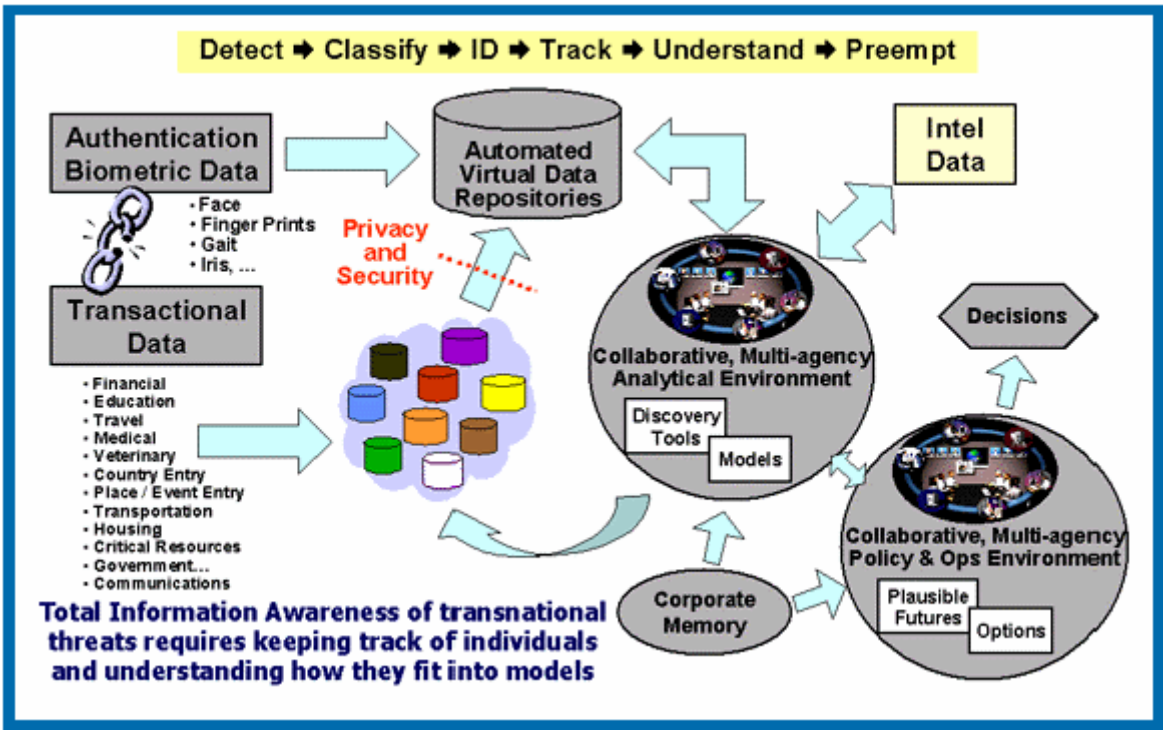

Figure 2. TIA chart of data sources and data flow (Stanley, 2003, p. 3)

Table 1

Federal Government Information	The Privacy Act of 1974
Education Information.	The Family Education Rights and Privacy Act of 1974
Telecommunications Information	The Cable Communications Policy Act of 1984 The Video Protection Act of 1988 Telecommunications Act of 1996
Health Information	The Health Insurance Portability and Accountability Act of 1996
Motor Vehicle Information	Driver's Privacy Protection Act of 1994
Communications and Communications Record	Title III of the Omnibus Crime Control and Safe Streets Act of 1968 The Foreign Intelligence Surveillance Act of 1978 The Electronic Communications Privacy Act of 1986 The USA PATRIOT Act of 2001 The Homeland Security Act of 2002
Financial Information.	The Fair Credit Reporting Act of 1970 The Right to Financial Privacy Act of 1978 The Gramm-Leach-Billey Act of 1999 The Bank Secrecy Act of 1970 The Currency and Foreign Transaction Reporting Act The Tax Reform Act of 1976 The Tax Equity and Fiscal Responsibility Act of 1982
Other Information.	Children's Online Privacy Protection Act of 1998 United States Constitution, Fourth Amendment Title 18 of the United States Code

Table 1. Statutory provisions related to TIA